DOCUMENT RESUME

ED 375 567 EC 303 404

AUTHOR	Thurlow, Martha L.
TITLE	National and State Perspectives on Performance Assessment and Students with Disabilities. Performance Assessment: CEC Mini-Library.
INSTITUTION	Council for Exceptional Children, Reston, Va.; ERIC Clearinghouse on Disabilities and Gifted Education, Reston, VA. ERIC/OSEP Special Project on Interagency Information Dissemination.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
REPORT NO	ISBN-0-86586-251-6
PUB DATE	94
CONTRACT	RR93002005
NOTE	53p.
AVAILABLE FROM	Council for Exceptional Children, 1920 Association Dr., Reston, VA 22091-1589 ($8.90 non-members, $6.25 members; Stock No. P5060; $32 set, non-members, $22.50 set, members; Stock No. P5062).
PUB TYPE	Information Analyses (070)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Standards; Data Collection; *Disabilities; *Educational Assessment; *Educational Change; Educational Legislation; Educational Objectives; Elementary Secondary Education; Evaluation Methods; *National Programs; *State Programs; *Student Evaluation
IDENTIFIERS	National Adult Literacy Survey (NCES); National Assessment of Educational Progress; *Performance Based Evaluation; Testing Accommodations (Disabilities)

ABSTRACT

This booklet examines national and state educational reform in the 1990s, noting the ways in which performance assessment is being presented as a mechanism of reform. Three reform initiatives are examined: national education goals, standards, and reform legislation. In conjunction with these reform movements, new methods of assessment are being called for. Large amounts of money and excitement have surrounded national and state reform activities that either directly or indirectly are connected to the idea of performance-based assessments. Major national data collection efforts have changed to adopt the performance assessment approach, including the National Assessment of Educational Progress and the National Adult Literacy Survey. Participation of students with disabilities in national assessments is problematic, constrained by the lack of accommodations. The use of performance assessment is also increasing in statewide assessment programs, requiring that accommodations and adaptations for students with disabilities be considered. There is some evidence that the use of performance assessments may not benefit students with disabilities, and to date, their use has not increased the participation of students with disabilities, but they hold the possibility of more equitable student measurement. (Contains 43 references.) (JDD)

ED 375 567

CEC Mini-Library

PERFORMANCE ASSESSMENT

National and State Perspectives on Performance Assessment and Students with Disabilities

Martha L. Thurlow

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Funded by the Office of Special Education Programs, U.S. Department of Education

EC 303404

BEST COPY AVAILABLE

The Council for Exceptional Children

CEC: Leading the Way

The Council for Exceptional Children (CEC) is the largest professional organization internationally committed to improving educational outcomes for individuals with exceptionalities. CEC accomplishes its worldwide mission on behalf of educators and others working with children with exceptionalities by advocating for appropriate government policies; setting professional standards; providing continuing professional development; and assisting professionals to obtain conditions and resources necessary for effective professional practice.

CEC: The Unifying Force of a Diverse Field

The Council for Exceptional Children, a private nonprofit membership organization, was established in 1922. CEC is an active network of 59 State/Provincial Federations, 900 Chapters, 17 Specialized Divisions, and 275 Subdivisions with reach in over 40 countries.

The CEC Information Center:
International Resource for Topics in Special and Gifted Education

The Council for Exceptional Children is a major publisher of special education literature and produces a comprehensive catalog semiannually. Journals such as *TEACHING Exceptional Children* (published quarterly) and *Exceptional Children* (published 6 times per year) reach over 100,000 readers and provide a wealth of information on the latest teaching strategies, research, resources, and special education news. *CEC Today*, the Council's monthly communications vehicle, provides the latest on Association issues, events, and advocacy efforts.

To help those in the field respond to their ever-increasing classroom challenges, CEC is proud to release *National and State Perspectives on Performance Assessment and Students with Disabilities.*

The Council for Exceptional Children
1920 Association Drive • Reston, Virginia 22091
703/620-3660 (Voice/TDD)
703/264-9494 (Fax)

CEC Mini-Library

PERFORMANCE ASSESSMENT

National and State Perspectives on Performance Assessment and Students with Disabilities

Martha L. Thurlow

Published by
The Council for Exceptional Children

ERIC

A Product of the
ERIC/OSEP Special Project
The ERIC Clearinghouse on
Disabilities and Gifted Education

Library of Congress Cataloging-in-Publication Data

Thurlow, Martha L.
National and State perspectives on performance assessment and students with disabilities / Martha L. Thurlow.
p. cm. — (CEC mini-library performance assessment)
"A product of the ERIC/OSEP Special Project, the ERIC Clearinghouse on Disabilities and Gifted Education."
Includes bibliographical references.
ISBN 0-86586-251-6
1. Educational tests and measurements—United States.
2. Handicapped students—United States—Rating of. 3. Educational accountability—United States. I. Title. II. Series.
LB3051.T52 1994
371.2'7—dc20 94-26213
CIP

ISBN 0-86586-251-6

A product of the ERIC/OSEP Special Project, the ERIC Clearinghouse on Disabilities and Gifted Education

Published in 1994 by The Council for Exceptional Children, 1920 Association Drive, Reston, Virginia 22091-1589

Stock No. P5060

This publication was prepared with funding from the U.S. Department of Education, Office of Special Education Programs, contract no. RR93002005. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication the manuscript was submitted for critical review and determination of professional competence. This publication has met such standards. Points of view, however, do not necessarily represent the official view or opinions of either The Council for Exceptional Children or the Department of Education.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Foreword

CEC's policy on inclusive schools and community settings invites all educators, other professionals, and family members to work together to create early intervention, educational, and vocational programs and experiences that are collegial, inclusive, and responsive to the diversity of children, youth, and young adults. Policymakers at the highest levels of state/provincial and local government, as well as school administration, also must support inclusive principles in the educational reforms they espouse.

One area in which the inclusion of students with disabilities is critical is the development and use of new forms of assessment. This is especially true when assessment becomes a tool by which local school districts, states, and our nation show accountability for the education of students.

As multidimensional instruments that can cross curriculum areas, performance assessments have the potential to be powerful instructional tools as well as tools for accountability. As this new technology is applied in creating new assessment instruments, students with disabilities must be considered during the design of the assessment, administration, scoring, and reporting of results.

CEC is proud to contribute this Mini-Library to the literature on performance assessment, and in so doing to foster the appropiate inclusion of students with disabilities in this emerging technology for instruction and accountability.

Preface

Performance assessment, authentic assessment, portfolio assessment—these are the watchwords of a new movement in educational testing. Its advocates say this movement is taking us beyond the era when the number 2 pencil was seen as an instrument of divine revelation. Its critics say it is just another educational bandwagon carrying a load of untested techniques and unrealistic expectations.

Despite the criticisms and reservations that are sometimes expressed, these new approaches are being implemented in a growing number of large-scale assessment programs at federal, state, and district levels. They are also finding their way into small-scale use at school and classroom levels.

What about students with disabilities? Are the new assessment techniques more valid than conventional assessment techniques for these students? Are the techniques reliable and technically sound? Will they help or hinder the inclusion of students with disabilities in large-scale assessment programs? Can classroom teachers use the techniques to assess student learning and possibly enrich the classroom curriculum?

The following fictional vignettes illustrate some of these issues.

Vignette 1

The State of Yorksylvania developed educational standards and a statewide system of student assessments to monitor progress in achieving the standards. The use of standardized multiple-choice tests was rejected because these tests were thought to trivialize education. It was feared that teachers would "teach down" to the tests rather than "teach up" to the standards. So, committees of teachers, parents, and employers were formed to translate the standards into "authentic" performance assessments. The resulting assessment system was called the Yorksylvania Performance Inventory (YPI).

Once a year, students from every school in the state were administered the YPI, which consisted of several assessments, each of which required up to 3 days to complete. Students worked, sometimes individually and sometimes in small groups, on tests involving complex, high-level tasks that crossed curriculum areas. In one task, students individually did research and answered essay questions interrelating the geography, wildlife, and history of their state. In another task, students worked in groups to design a car powered by fermentation. Schools were provided with practice activities and curriculum guides to encourage the infusion of performance assessment activities into the school curriculum.

The state policy allowed special education students to be included in the YPI, excluded, or provided with special modifications, depending on their individual needs as indicated in their individualized education programs. Initially, most special education teachers supported the YPI because they felt it eliminated some artificial barriers (reading, test-taking skills, etc.) that put their students at a disadvantage on other types of tests. However, there were some questions and issues, such as the following:

- Some of the YPI tasks involved *a lot* of reading, more than was found on previous types of tests.
- Special education teachers sometimes felt pressured to exclude their students from testing in order to increase the school's scores.
- Special education students sometimes experienced extreme frustration in the YPI assessments, many of which bore no resemblance to these students' other schoolwork.
- Some parents of special education students questioned whether the standards were really applicable to their children and whether the YPI was diverting instruction from more relevant and important topics.

Vignette 2

A teacher named Pat had students at a wide range of functioning levels, including a number of mainstreamed students receiving special education services. Pat was always on the lookout for new ideas and approaches. Pat began reading articles and attending conferences on new assessment approaches termed *portfolio assessment, authentic assessment, per-*

formance assessment, and *alternative assessment*. These approaches seemed to make a lot of sense, and Pat decided to try them out. One of the first approaches Pat tried was authentic assessment. Rather than simply testing students on their rote learning of skills and content, Pat began to look for ways to use realistic, complex activities to test whether the students could actually apply what they learned. For example, Pat combined writing, spelling, science, and career skills into an activity in which students wrote letters of application for jobs as physicists, biologists, or chemists. Pat particularly valued activities that engaged students in solving interesting problems. For example, after a unit on optics, Pat assigned students to draw a diagram explaining why mirrors reverse an image from left to right but not from top to bottom. The students grappled with that problem for several days.

Pat liked the holistic scoring procedures developed in these new assessment approaches. Rather than simply marking a response correct or incorrect, Pat scored student work on a number of dimensions (e.g., analysis of the problem, clarity of communication) according to meaningful quality criteria. The development of authentic performance tasks and scoring procedures helped Pat clarify the most important learning outcomes.

Pat also liked the idea of portfolio assessment, in which students could select and collect "best pieces" to demonstrate their learning and achievement during the year. Student self-evaluation became a valued part of this process.

In all, Pat was very pleased with these new assessment approaches and intended to continue using them. Instruction became more activity based and more focused on real-world uses of the material. There were, however, some issues that Pat began to think about:

- Students with deficits in certain academic areas, notably writing, were at a real disadvantage. It was sometimes hard to determine whether an inadequate response resulted from poor writing skills, poor mastery of the content, poor problem-solving skills, lack of creativity, or some combination of these factors. Pat considered allowing some students to tape record their responses, but decided not to. Wasn't writing itself an authentic task required in the real world?

- Pat wasn't sure how to use the information provided by these tests to plan additional instruction, particularly if a student was having difficulty.
- Pat wondered how to tell whether or not an activity was in fact authentic, especially for students whose adult lives would be very different from Pat's own.

In 1992, the Division of Innovation and Development (DID) in the U.S. Department of Education's Office of Special Education Programs and the ERIC/OSEP Special Project of The Council for Exceptional Children formed a Performance Assessment Working Group to discuss issues such as these. The term *performance assessment* was adopted as a general designation for the range of approaches that include performance assessment, authentic assessment, alternative assessment, and portfolio assessment.

Performance assessment was defined has having the following characteristics:

1. *The student is required to create an answer or a product rather than simply fill in a blank, select a correct answer from a list, or decide whether a statement is true or false.*
2. *The tasks are intended to be "authentic."* The conventional approach to test development involves selecting items that represent curricular areas or theoretical constructs, and that have desired technical characteristics (e.g. they correlated with other similar items, they discriminated between groups, etc.). Authentic tasks, on the other hand, are selected because they are "valued in their own right"[1] rather than being "proxies or estimators of actual learning goals."[2]

The Performance Assessment Working Group produced this series of four Mini-Library books on various topics related to performance assessment and students with disabilities. In *National and State Perspectives on Performance Assessment and Students with Disabilities*, Martha Thurlow discusses trends in the use of performance assessment in large-scale testing programs. In *Performance Assessment and Students with Disabilities: Usage in Outcomes-Based Accountability Systems*, Margaret McLaughlin and Sandra Hopfengardner Warren describe the experi-

[1] R. L. Linn, E. L. Baker, & S. B. Dunbar. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.
[2] M. W. Kirst. (1991). Interview on assessment issues with Lorrie Shepard. *Educational Researcher, 20*(2), 21–23, 27.

ences of state and local school districts in implementing performance assessment. In *Creating Meaningful Performance Assessments: Fundamental Concepts*, Stephen Elliott discusses some of the key technical issues involved in the use of performance assessment. And, in *Connecting Performance Assessment to Instruction*, Lynn Fuchs discusses the classroom use of performance assessment by teachers.

Martha J. Coutinho
University of Central Florida

David B. Malouf
U.S. Office of Special Education Programs

August, 1994

Members of the Performance Assessment Work Group

Joan Baron, Performance Assessment Collaborative for Education, Harvard University

Joyce Choate, Council for Exceptional Diagnostic Services, Northeast Louisiana University

Lorraine Costella, Frederick County, Maryland, Public Schools

Martha Coutinho, Division of Innovation and Development, U.S. Office of Special Education Programs

Stephen Elliott, University of Wisconsin-Madison

Lynn Fuchs, Vanderbilt University

John Haigh, Maryland State Department of Education

Larry Irvin, University of Oregon

Robert Linn, Center for Research on Evaluation, Standards, and Student Testing

Lynn Malarz, Association for Supervision and Curriculum Development

David Malouf, Division of Innovation and Development, U.S. Office of Special Education Programs

Margaret McLaughlin, Center for Policy Options in Special Education, University of Maryland

Trina Osher, National Association of State Directors of Special Education

James Poteet, Council for Educational Diagnostic Services, Ball State University

Clay Starlin, Western Regional Resource Center, University of Oregon

Martha Thurlow, National Center on Educational Outcomes for Students with Disabilities, University of Minnesota

Sandra Hopfengardner Warren, Center for Policy Options in Special Education, University of Maryland

Note: Members' affiliations may have changed since the work group was formed.

About the Author

Martha L. Thurlow is Assistant Director at the National Center on Educational Outcomes for Students with Disabilities (NCEO), which is located at the University of Minnesota. She currently is also a co-principal investigator of a dropout prevention project for middle-school-age students with learning and emotional disabilities. In the past, her research has focused on school-based models of mainstreaming, the follow-up of students with disabilities leaving school, early childhood assessment, and students with learning disabilities.

"Educational policies are too often set without any consideration of their impact on students with disabilities," says Dr. Thurlow. "We need to be proactive in considering the consequences of educational reforms for students with disabilities. Ideally, we should document and examine the unintended consequences of policies and resulting practices." The need for relevant research for practitioners and policy makers has been a driving force for much of Dr. Thurlow's research.

Dr. Thurlow has been associated with many research institutes and centers during the past 24 years. Her interest in individuals with disabilities was first sparked during high school volunteer work at institutional settings for individuals with mental retardation. Her research interests over the years have broadened in scope from her start in basic research to descriptive research, and then to intervention research, and now policy research.

Dr. Thurlow recently authored, with James Ysseldyke and Bob Algozzine, a book entitled *Critical Issues in Special Education*. She is currently working on other books that will examine in greater depth the implications of educational reform legislation for students with disabilities.

Contents

1. **Introduction, 1**

2. **National and State Education Reform in the 1990s, 3**

   Major Reform Efforts, 3
   Emergence of Assessment as a Mechanism of Reform, 8
   Calls for New Methods of Assessment, 9

3. **National Data-Collection Programs, 15**

   Performance Assessments Within National Programs, 15
   Special Education in National Assessments, 20

4. **State Data-Collection Programs, 23**

   State Efforts, 23
   Special Education in Statewide Performance Assessments, 25
   Accommodations and Adaptations in Statewide Performance Assessments, 26

5. **Conclusions, 30**

References, 34

1. Introduction

Assessment has always been an important part of education. It has served a variety of purposes, from measuring student progress to describing the national condition of education. Different forms of assessment have been used for different purposes. This practice reflects the wisdom of most professionals, who agree on the need for differentiation (e.g., Haney, 1991).

The emphasis on educational reform in the past decade included the assessments used to measure progress. Large-scale assessments, typically used to describe the educational status of a large group of students, were viewed from a new perspective. The information obtained from them was seen as more important than ever before, because it revealed the status of the nation or a state in achieving educational goals. As the emphasis on assessment increased, there was a corresponding increase in concern about the adequacy of the most common form of assessment being used for large-scale assessments, the traditional multiple-choice test (e.g., Cannell, 1988; Linn, Baker, & Dunbar, 1991). Policymakers noted, for example, that even those students "who succeed in school and score well on conventional tests have not been educated to cope successfully with the demands of personal, vocational, and civic life in contemporary society" (Newmann, 1991, p. 459). In other words, the assessments were not measuring what needed to be measured to ensure that those who performed well on the test also would perform well in society.

At both the national and state levels, there is now a flurry of activities under way to rethink and reframe large-scale assessment systems. These activities are pointing toward greater use of performance assessments in large-scale assessment programs. *Performance assessment* is one of many terms coined for the new type of assessment that would enable students to demonstrate their "authentic" knowledge, that is, skills and content that are meaningful and motivational to the student and that are related to functioning in the world beyond the school walls. The term is used to describe assessments that "require students to create an answer or product that demonstrates their knowledge or skills"

(Office of Technology Assessment [OTA], 1992, p. 3). Activities such as open-ended writing, making a presentation, and preparing a portfolio of student work all would be considered performance assessments.

While performance-based assessments have been used for some time to make instructional decisions for individual children, their use in large-scale assessments as a way to monitor the educational system is relatively new. Large-scale assessments are those that produce information about large numbers of students, thereby making it possible to summarize the status of education on a broad scale and to conduct subgroup analyses (e.g., comparisons of progress of students from different cultural backgrounds). Large-scale assessments typically are used to monitor the educational system (OTA, 1992).

Several issues emerge for students receiving special education services when performance-based items are used in large-scale assessments. To better understand what this emphasis may mean for students with disabilities, it is important to have a grasp of what is happening at both the national and state levels. This book examines national and state educational reform in the 1990s, noting the ways in which performance assessment is being presented as a mechanism of reform. Major national data-collection efforts that have changed to adopt the performance assessment approach, in part or in whole, are explored. Finally, information is provided on the use of performance assessment in statewide assessment programs. For each of these topics, the implications for students with disabilities are examined.

2. National and State Education Reform in the 1990s

Education reform efforts mushroomed in the late 1980s and early 1990s. Following waves of concern about education that arose in the early 1980s, reform efforts came to the forefront when a set of national education goals was defined, higher standards were promoted, and education reform legislation was enacted. To better understand national and state education reform efforts of the 1990s, it is helpful to look at the major reform efforts taking place, the emergence of assessment as a mechanism of reform, and calls for new methods of assessment.

Major Reform Efforts

At least three recent reform initiatives reflect the emphasis on assessment and the trend toward viewing performance assessment as a part of large-scale assessments: national education goals, standards, and reform legislation.

National Education Goals

In the fall of 1989, President Bush and the governors held an education summit. It was at this meeting that six national education goals to be reached by the year 2000 were established:

1. Every child in the United States will start school ready to learn.
2. The high school graduation rate will reach 90%.
3. Every American student will achieve competence in challenging subject matter, including English, mathematics, science, history, and geography; and every school will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment.
4. The United States will be first in the world in science and mathematics.

5. Every American adult will be literate and a life-long learner.
6. All American schools will be safe, disciplined, and drug-free environments in which all students will be able to learn.

The impact of these goals is reflected in nearly all states in the setting of state-level education goals that correspond closely to the national education goals (see Figure 1). The fact that one of the leading governors working on the identification of the goals (Clinton) is now president has ensured that emphasis continues to be placed on the importance of reaching these goals by the year 2000. These goals continued to take precedence, even though two goals were added and one goal was changed when they were codified through the passage of Goals 2000 legislation.

The fact that one of the leading governors working on the identification of the goals (Clinton) is now president has ensured that emphasis continues to be placed on the importance of reaching these goals by the year 2000.

One of the first challenges that accompanied the setting of goals was the need to identify ways to measure progress toward achieving the goals. The National Education Goals Panel (NEGP) was established to carry out this task. In its efforts to do this, NEGP formed task forces and work groups to focus on measurement issues. The National Council on Education Standards and Testing (NCEST) was formed to address the setting of standards and the development of assessments for a set of core academic subject areas (English, mathematics, science, history, and geography).

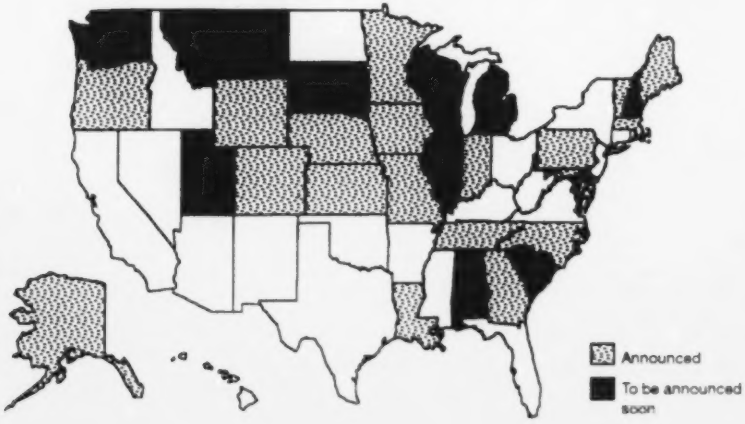
Standards

The notion of setting higher standards to ensure that students would become competent in core academic areas was proposed by a Goal 3 work group. NCEST studied the feasibility of setting standards and assessing progress toward them. In January 1992, NCEST produced *Raising Standards for American Education*, in which it argued that higher standards were needed for *all* American students. NCEST proposed that these standards should challenge the most able students as well as those with special learning needs:

> The Council's intent in recommending the establishment of national standards is to raise the ceiling for students who are

FIGURE 1
States with Goals That Correspond Closely to the Original Six National Education Goals

AMERICA 2000: SIX MONTHS LATER
October, 1991



AMERICA 2000: ONE YEAR LATER
March, 1992

(*Source*: U.S. Department of Education, *America 2000*, Numbers 7 & 23).

> currently above average and to lift the floor for those who now experience the least success in school, including those with special needs. (National Council on Education Standards and Testing [NCEST], 1992, p. 4)

Inextricably linked with standards, in the view of NCEST, is assessment. Throughout the *Raising Standards* document are references to "national standards and a system of assessments." The rationale behind this linkage was explained in the document:

> The Council determined that it is not sufficient just to set standards. Since tests tend to influence what is taught, assessments should be developed that embody the new high standards. The considerable resources and effort the Nation expends on the current patchwork of tests should be redirected toward the development of a new system of assessments. Assessments should be state-of-the-art, building on the best tests available and incorporating new methods. (p. 4)

While not specifically endorsing the use of performance assessment at the national level, the *Raising Standards* document was one of the first to specifically discuss the possibility of using such assessments:

> There is significant interest in the promise of performance-based assessments, such as portfolios and projects, as ways of collecting evidence of what students know and can do. Such assessments frequently use open-ended tasks, focus on higher-order or complex thinking skills, require significant student time, and may allow students to choose among alternative tasks; some examine the performance of group activities. While important issues remain to be resolved, innovative techniques used by states and localities may be important elements in the mix of assessment instruments that will make up the new national system. (p. 28)

The main concern of NCEST was that safeguards be built in to the system "to protect students from negative consequences while the system of assessments is being refined, especially for students who have not been well served by testing in the past" (pp. 29–30).

The New Standards Project, which was established in the early 1990s, focused on both standards and assessments. This project represented the collaborative work of the Learning Research and Development Center (LRDC) at the University of Pittsburgh and the National

Center for Education and the Economy (NCEE). The purpose of the New Standards Project was to establish a national examination system. As proposed, this system would consist of three components: a performance examination, assessments of student projects, and assessments of the contents of a portfolio of student work. The performance examination is supposed to assess mastery of bodies of knowledge and focus on thinking, problem solving, and application of knowledge to real-life problems. The New Standards Project is developing a reference examination and guiding participating states in the development of their own assessments, with the notion that states should develop assessments that best meet their local needs and that a linking system can be developed to pull the results of individual states together to produce "national examination" results.

To date, 18 states and 6 districts are working in partnership with the New Standards Project to develop national standards and performance-based assessments in math, science, English, language arts, and history. Fourth-grade English and math assessments have been pilot tested.

Reform Legislation

The importance of the six national education goals, setting standards in core academic areas, and assessment is reflected in the Goals 2000: Educate America Act, the education reform legislation proposed by President Clinton and signed by him on March 31, 1993. The final law has 10 titles that encompass and expand upon the original 4 titles. The first 5 titles of Goals 2000 are as follows:

Title I. National Education Goals

Codifies in law the original six national education goals—with Goal 3 expanded to include civics, economics, and the arts—and adds two new goals, one on teacher training (new Goal 4), and one on parent involvement (new Goal 8).

Title II. National Education Reform Leadership, Standards and Assessment

Establishes in law the National Education Goals Panel (NEGP), which oversees progress toward goals and establishes a National Education Standards and Improvement Council (NESIC) to approve national voluntary standards and state proposed standards.

Title III. State and Local Education Systemic Improvement

Supports statewide and local reform efforts through a

state grant program. To receive funds, a state must establish a State Planning Panel, which will develop a comprehensive reform plan. This plan is to identify strategies for

- developing or adopting standards
- providing students the opportunity to learn
- management and governance to promote accountability
- involving parents and the community
- bringing education reform to scale
- strategies for assisting local education agencies and schools to meet the needs of students who have dropped out of school

Title IV. Parental Assistance

Establishes a new discretionary grants program to promote parent information and participation in their child's education.

Title V. National Skills Standards Board

Establishes and funds a national board to develop job skills standards.

(Adapted from Shriner, Ysseldyke, & Thurlow, 1994, p. 15)

Assessment and the notion of using new forms of assessment are integral parts of the proposed legislation. In addition, these themes are being carried into other education legislation, thereby reinforcing the notion that different education programs are integrated and will be held accountable for results in similar ways. Specifically, the Elementary and Secondary Education Act (ESEA), which provides funding for Chapter I programs, and the Individuals with Disabilities Education Act (IDEA), which provides funding for special education programs, were reshaped within the context of the Goals 2000 legislation.

Emergence of Assessment as a Mechanism of Reform

With all the reform efforts under way, and so many of them focusing on assessments, it is no wonder that assessment has been viewed as a mechanism of reform. In its report to inform federal policymakers about testing (OTA, 1992), the Office of Technology Assessment noted that Americans have regarded standardized tests as multipurpose tools, with

one of the purposes being "agent of school reform." When NCEST published *Raising Standards*, it recognized that the development of standards and assessments of them could serve as a mechanism of reform:

> Developing standards and assessments at the national level can contribute to educational renewal in several ways. This effort has the potential to raise learning expectations at all levels of education, better target human and fiscal resources for educational improvement, and help meet the needs of an increasingly mobile population. Finally, standards and assessments linked to the standards can become the cornerstone of the fundamental systemic reform necessary to improve schools. (1992, p. 5)

Despite concerns about tests dictating what is taught, it has been recognized for some time that instruction and assessment are linked, as are reform efforts and testing. Major education reform efforts almost always require either expansion of existing testing or the development of new forms of testing (Pipho, 1985). The role of assessment changes when it is part of the reform it evaluates. It is because of the link between testing and instruction, in part, that there have been calls for new methods of assessment.

Calls for New Methods of Assessment

Among the first to call for new forms of assessment in large-scale assessment programs was Wiggins (1989). He argued, first, that we must "test those capacities and habits we think are essential, and test them in context" (p. 41), and second, that it is possible to use "authentic" tests on a large-scale basis. He suggested that "the supposed impracticality and/or expense of designing such tests on a wide scale is a habit of thinking, not a fact" (p. 44). Basically, the calls for new forms of assessment cried for a halt to assessment practices that reduced teaching to preparation for testing, narrowed the curriculum to areas tested, and focused instruction on simple skills rather than higher-order thinking (Berlak et al., 1992; Haney & Madaus, 1989; Moss, 1992; National Commission on Testing and Public Policy, 1990). Marzano, Pickering, and McTighe (1993) portrayed the move to performance assessment as a "revolution in assessment" that was needed to reflect broader educational goals and enhance learning and teaching. Performance assessment also was viewed as meeting the need for an improved record-keeping and reporting system.

The calls for new forms of assessment cried for a halt to assessment practices that reduced teaching to preparation for testing, narrowed the curriculum to areas tested, and focused instruction on simple skills rather than higher-order thinking.

Newmann (1991) listed several reasons for moving toward authentic assessments: "Participation in authentic tasks is more likely to motivate students . . . students will have a greater stake in authentic achievement . . . authentic academic challenges are more likely to cultivate the higher-order thinking and problem-solving capacities" (p. 460). Even though the calls for new forms of assessment might not be considered new, the emphasis given to their use in large-scale assessments and for accountability purposes was (Mehrens, 1992).

Along with the calls for new methods of assessment have come a blurring of the meaning of terms describing assessment. Hill and Larsen (1992) referred to the instability of the term *testing* in describing assessment activities. They noted also that many test makers are claiming that their tests are authentic and test higher-order thinking skills even though they still use multiple-choice items. This has been a frequent phenomenon because test publishers almost immediately began to claim that their assessments were "performance based and authentic." Hill and Larsen cautioned that "teachers and administrators at all levels need to be wary of the claims that accompany multiple-choice tests. As test makers rush to join the movement for greater authenticity in assessment, they often end up constructing a test that is more dysfunctional than a conventional one (p. 23)."

Perhaps the clearest examples of what new methods of assessment were being lauded were those presented by the groups involved in setting standards in various content areas. Almost immediately after the *Raising Standards* document was released, groups sprang up to develop standards in key content areas. Several of these groups (geography, history, civics, science, English, foreign languages, arts) were funded, in part or in whole, by the U. S. Department of Education's Office of Educational Research and Improvement (OERI). One standards-setting group, the National Council of Teachers of Mathematics (NCTM), had set standards even before the six national educational goals were identified and NCEST was formed. In fact, NCEST repeatedly referred to the work of NCTM in its *Raising Standards* document.

The NCTM (1989) document *Curriculum and Evaluation Standards for School Mathematics* presented four cornerstone standards (mathemat-

ics as problem solving, communication, reasoning, and connections) plus additional standards for three grade groups (K–4, 5–8, 9–12). Two years later, an *Addenda Series* (Burton et al., 1991) was published to provide assistance to teachers in implementing instruction to support the NCTM standards. NCTM (1993) is now working on its assessment publication to support the standards. This draft document emphasizes the importance of assessment as the intersection of teaching and learning and as a means for supporting the learning of each student to develop "mathematical power" in *all* students. In this document, NCTM presents six assessment standards to judge the appropriateness of assessments:

1. *Important mathematics.* Assessment should reflect the mathematics that is most important for students to learn.
2. *Enhanced learning.* Assessment should enhance mathematics learning.
3. *Equity.* Assessment should promote equity by giving each student optimal opportunities to demonstrate mathematical power and by helping each student meet the profession's high expectations.
4. *Openness.* All aspects of the mathematics assessment process should be open to review and scrutiny.
5. *Valid inferences.* Evidence from assessment activities should yield valid inferences about students' mathematics learning.
6. *Consistency.* Every aspect of an assessment should be consistent with the purposes of the assessment.

NCTM believes that these standards are ones that could form the basis for developing "new effective assessment systems," and that "current commonly used assessment instruments (norm-referenced standardized tests, textbook tests, state and national profile examinations) and inferences based on their use would fail miserably when judged against these standards" (p. 2).

The efforts of the group setting standards for science also illustrate the nature of the new methods of assessment seen as congruent with higher standards and higher-order problem solving and thinking. A comprehensive set of content, teaching, and assessment standards is being prepared by the National Committee on Science Education Standards and Assessment (NCSESA). Just as the science content standards are organized into major areas (science as inquiry, science subject matter, scientific connections, scientific and human affairs), the assessment standards are in five areas (assessment in the service of learning from the student's perspective; assessment in the service of teaching and

learning from the teacher's perspective; assessment for decisions about individuals; assessment for policy; assessment to monitor the system). These standards (see National Research Council, 1993) are currently being completed, and a sample assessment is being prepared.

In addition to the standards-setting groups, the Office of Educational Research and Improvement (OERI) also funded the National Center for Research on Evaluation, Standards, and Student Testing (NCRESST). NCRESST has addressed the issues surrounding performance assessments with great gusto and much writing. Some of the documents it has produced within the past few years are listed in Table 1. As is evident from this list, the topics cover everything from existing uses of performance assessments to research on questions about the technical adequacy of such measures.

Additional evidence of the recognition of performance assessment as the favored new method of assessment is the production of numerous documents and even videotapes on the topic (e.g., a videotape, *Alternatives for Measuring Performance*, produced by the North Central Regional Education Laboratory and the Center for Research on Educational Standards and Student Testing [CRESST]). CRESST also supports an Internet server called *Alternative Assessments in Practice Database*, which contains source information on more than 250 alternative assessments currently in use. According to a brochure produced by CRESST, the database provides easy access to information about ongoing and newly developed measures from states, curriculum and teacher groups, and other research and development sources. Subjects targeted by the assessments summarized in the database include language arts, mathematics, science, social studies, foreign language, workforce readiness, and fine arts.

Money and excitement have surrounded a number of national and state reform activities that either directly or indirectly are connected to the idea of performance-based assessments.

The many activities surrounding assessment, most of which were responding to calls for new forms of assessment, reflect the attention paid to performance-based assessment as a part of educational reform. Moreover, the activities were backed by considerable funds, a sure signal of their importance. Large sums of money also were being directed toward efforts either to study performance assessment or to develop performance assessment measures. Some of the funding figures are presented in Table 2. As noted in the table, for example, the New

---

TABLE 1
Selected NCRESST Publications on Performance Assessment

CSE Technical Report Number (Year of Publication)		*Publication Title*
331	(1991)	*Complex, Performance-Based Assessment: Expectations and Validation Criteria*
335	(1992)	*Cross-State Comparability of Judgments of Student Writing: Results from the New Standards Project*
337	(1992)	*Writing Portfolios at the Elementary Level: A Study of Methods for Writing Assessment*
341	(1991)	*Implications for Diversity in Human Characteristics for Authentic Assessment*
348	(1992)	*Accountability and Alternative Assessment*
349	(1992)	*Design Characteristics of Science Performance Assessments*
350	(1992)	*The Vermont Portfolio Assessment Program: Interim Report on Implementation and Impact, 1991–92 School Year*
355	(1993)	*The Reliability of Scores from the 1992 Vermont Portfolio Assessment Program*
361	(1993)	*Sampling Variability of Performance Assessments*
362	(1993)	*Performance-Based Assessment and What Teachers Need*

Standards Project is working with funding of approximately $11 million. NCRESST was funded for $14 million for 5 years. Individual states are expending fairly substantial sums of money investigating or using performance assessments.

Thus, money and excitement have surrounded a number of national and state reform activities that either directly or indirectly are connected to the idea of performance-based assessments. Although the setting of national goals, the discussion about higher standards and assessments of them, and the legislative activities were among the most evident examples of the push toward performance-based assessment,

TABLE 2
Funding of Projects Dealing with Performance-Based Assessments

Project	*Funds Provided*
New Standards Project	$11 million
OERI Standards Projects*	
Arts	$1 million (2 years)
Civics	$780,000 (2 years)
English	$1.8 million (3 years)
Foreign Languages	$212,000 (1st of 3 years)
Geography	$700,000 (1 year)
History	$1.6 million (3 years)
Science	$3 million (2 years)
NCRESST	$14 million (5 years)

* Not all funds are provided by OERI. Other foundations and agencies also are providing funds.

there have also been a number of specific efforts to use performance-based assessment. These efforts, which include national data-collection programs and statewide assessments, illustrate some of the implications of these assessments for students with disabilities.

3. National Data-Collection Programs

Tests are used in classrooms throughout the country every day. Many people do not realize, however, that the United States has a comprehensive assessment program at the federal level as well. Even though the United States is one of the few countries without a national examination, it does collect a tremendous amount of information on its students. This chapter describes the use of performance assessment items in some of our existing national data-collection programs. The role of special education in these efforts is explored, specifically in terms of the participation of students with disabilities in national assessment.

Performance Assessments Within National Programs

National data-collection programs in education typically have relied on standard multiple-choice, paper-and-pencil exams to assess the status of American education. Although there were some cases in which performance-based assessments were used in the past, assessment programs that are relatively recent in origin are more likely to incorporate what might be called performance assessment items. Two relevant national education data-collection programs are the National Assessment of Educational Progress (NAEP) and the National Adult Literacy Survey (NALS).

National Assessment of Educational Progress

NAEP is known as our nation's "report card" and is considered to be the primary survey of educational achievement of American students and changes in achievement across time. It was initiated in 1969 to assess achievement of national samples of students in core subject areas. Typically, one content area was assessed every other year and data were reported only for the nation as a whole and for regions of the country. With the escalating interest in monitoring the achievement of American students, the number of subject areas assessed and the frequency of administration have increased. Also, in 1990 a voluntary NAEP trial state

assessment program was started to determine the feasibility and value of providing information at the state level, so that state and local policies could be linked to achievement data.

Traditionally, NAEP was an objective, multiple-choice, paper-and-pencil test. Over time, however, NAEP has responded to changing perspectives on achievement assessment, adapting with changes in content focus and types of items. Thus, in the past, NAEP had used some items that could be called "performance based." With the recent emphasis on authentic, performance-based achievement information, NAEP again has added items to its assessments that are performance based. A recent summary of NAEP initiatives by the Education Commission of the States (1992) indicated that innovations in NAEP included "assessing math performance with and without calculators; using open-ended items; assessing higher-order thinking skills; portfolio assessments; [and] oral reading assessments" (p. 9).

In its 1992 assessments, NAEP tested in the areas of reading and mathematics. As a result of an intensive assessment framework development process, three purposes for reading were identified (to gain literary experience, to gain information, and to perform a task) and crossed with types of interactions with text (initial understanding, developing an interpretation, personal reflection and response, and demonstrating a critical stance) (National Center for Education Statistics [NCES], 1993d). Both multiple-choice and constructed-response formats were used in this assessment, with "approximately 60 to 70 percent of the students' response time . . . devoted to constructed response questions" (NCES, 1993d, p. 44).

Two types of constructed-response items were used in the 1992 NAEP reading assessment. One type (regular constructed response) required short answers, from a few words to a few sentences. These were rated as either satisfactory or unsatisfactory. The second type (extended constructed response) required longer answers of a paragraph or more. These were rated on a four-point scale from unsatisfactory to extensive. Each reading passage presented in the assessment had at least one extended constructed-response question. Figure 2 provides examples of these "performance-based" items.

In mathematics, the 1992 NAEP used a framework that included five content areas (numbers and operations; measurement; geometry; data analysis, statistics, and probability; and algebra and functions) and three math abilities (conceptual understanding, procedural knowledge, and problem solving) (NCES, 1993b). In addition, NAEP was developed to be consistent with the NCTM standards.

Both multiple-choice and constructed-response formats (regular and extended) were used (NCES, 1993b). In addition, students were

FIGURE 2
Examples of NAEP Reading Performance-Based Items

Regular Constructed-Response Item:

Grade 4 (Student reads an informative article about how Amanda Clement became the first paid woman umpire.) **Write a paragraph explaining how Mandy got her first chance to be an umpire at a public game.**

_____
_____
_____

Grade 8 (Student reads and uses an actual bus schedule that includes tables, maps, and text.) **Monthly bus passes are not valid on which routes?**

_____
_____
_____

Extended Constructed-Response Item:

Grade 8 (Student reads two passages from the Oregon Trail, one an informational account of the Trail and the other a narrative piece based on a diary entry.) **Pretend that you are a young adult of the 1840s who has caught a case of "Oregon fever." Use information from both the passages and from your own knowledge to explain what you would do about Oregon fever and why.**

_____
_____
_____
_____
_____

Grade 12 (Student reads and uses an actual bus schedule that includes tables, maps, and text.) **Now that you have looked carefully at the bus schedule, use your notes and make suggestions to help New Jersey Transit improve this schedule.**

_____
_____
_____
_____
_____

required to provide responses using protractors/rulers, calculators, and manipulable geometric shapes. Students were given 5 minutes to demonstrate (in writing or through diagrams) their mathematical reasoning and problem-solving ability. They were also "led by audiotape through a series of tasks designed to measure their estimation skills" (NCES, 1993c, p. 37). Examples of some of the constructed-response format items are presented in Figure 3.

NAEP Arts Assessment

The focus of the standards, and thus the NAEP arts assessment, includes dance, music, theatre, and the visual arts (including design, architecture, and the media arts).

The Council of Chief State School Officers (CCSSO, 1993a) is now drafting an NAEP arts education assessment framework. The assessment is proposed for administration in 1996. The initial framework is grounded in the standards that are being developed in arts education through a joint effort of the Music Educators National Conference, the American Alliance for Theatre and Education, the National Art Education Association, and the National Dance Association. The focus of the standards, and thus the NAEP arts assessment, includes dance, music, theatre, and the visual arts (including design, architecture, and the media arts). The assessment will include three types of processes that are common to all art areas being assessed (creating, performing and interpreting, and responding) and the application of two kinds of content (knowledge about the arts and technical, perceptual, intellectual, and expressive skills). In addressing the "how" of the arts assessment, consideration is being given to (a) the authenticity of tasks (they should be as close as possible to the genuine artistic behaviors); (b) the demand characteristics of tasks (they should elicit higher-order thinking); and (c) the response modalities that are tapped (appropriate aural and visual responses need to be developed). In general, the goal is to use performance tasks and to draw on a wide range of formats, possibly including portfolios, performance assessments (e.g., playing an instrument in a concert), observation, interviews, questionnaires, self-evaluations, and paper-and-pencil tasks.

National Adult Literacy Survey (NALS)

The development of a national survey of the literacy skills of U.S. citizens was initiated in 1988. Prior to this survey, the literacy of young adults and job seekers had been studied, but no study had been conducted of

FIGURE 3
Examples of NAEP Mathematics Performance-Based Items

Grade 4

Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

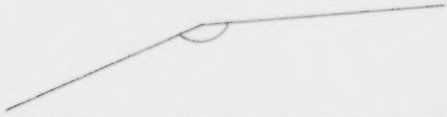
Laura wanted to enter the number 8375 into her calculator. By mistake, she entered the number 8275. Without clearing the calculator, how could she correct her mistake?

Without clearing the calculator, how could she correct her mistake another way?

Did you use the calculator on this question?

Yes No

Grade 8



Use your protractor to find the degree measure of the angle shown above.

Answer: _____

Grade 12

This question requires you to show your work and explain your reasoning. You may use drawings, words, and numbers in your explanation. Your answer should be clear enough so that another person could read it and understand your thinking. It is important that you show all your work.

One plan for a state income tax requires those persons with income of $10,000 or less to pay no tax and those persons with income greater than $10,000 to pay a tax of 6 percent only on the part of their income that exceeds $10,000.

A person's effective tax rate is defined as the percent of total income that is paid in tax.

Based on this definition, could any person's effective tax rate be 5 percent? Could it be 6 percent? Explain your answer. Include examples if necessary to justify your conclusions.

Did you use the calculator on this question?

Yes No

the general U.S. population. Based on the previous surveys and a framework of literacy skills that included prose, document, and quantitative literacy, a new set of literacy tasks was developed for the 1992 household interview survey. The following goals guided the development of new tasks:

- continued use of open-ended simulation tasks
- continued emphasis on tasks that measure a broad range of information-processing skills and cover a wide variety of contexts
- increased emphasis on simulation tasks that require brief written and/or oral responses
- increased emphasis on tasks that ask respondents to describe how they would set up and solve a problem
- the use of a simple, four-function calculator to solve selected quantitative problems

(NCES, 1993a, p. 4)

The literacy tasks that were built involved materials that "adults encounter in their daily activities" (p. 70). Prose materials included expository works, narratives, and poetry. Document materials included a variety of structures such as charts, tables, maps, and schedules. Quantitative materials involved numbers embedded within text.

Special Education in National Assessments

For the 1990 NAEP, both the national and the state trial assessment, approximately 45% to 50% of students with disabilities were excluded.

The participation of students with disabilities in national assessments has been fairly dismal (McGrew, Thurlow, & Spiegel, 1993). For the 1990 NAEP, both the national and the state trial assessment, approximately 45% to 50% of students with disabilities were excluded. The formal guidelines used by NAEP should not produce such high exclusion rates. The guidelines indicate that students who have individualized education programs (IEPs) may be excluded if "the student is mainstreamed less than 50 percent of the time in academic subjects and is judged to be incapable of taking part in the assessment or the IEP team has deter-

mined that the student is incapable of taking part meaningfully in the assessment" (Mullis, 1990, p. 36).

Difficulties with the implementation of NAEP guidelines emerged with full force when the trial state assessments were conducted. Exclusion rates ranging from 33% to 87% were found among the states participating in the assessment (McGrew et al., 1993). The conclusion of the National Academy of Education (1992), which studied the NAEP trial state assessment (TSA), was

> Significant variations in the state-by-state exclusions of IEP students were observed in the 1990 TSA that cannot be easily explained. Importantly, the Panel's research shows that differential exclusion rates affect the rankings of the states. Therefore the Panel recommends that NCES conduct a study designed to evaluate the rationales used by educators for the exclusion of IEP students on the basis of their ability to participate meaningfully in the assessment. This study would result in a better understanding of the differential use of exclusion criteria across states, thereby providing information that would allow states to compare themselves more accurately on NAEP assessments. (p. 13)

Issues related to whom to include in national assessments also emerged in NALS. During the field testing of the assessment, interviewers had skipped houses in which the person answering the door was unable to read. Fortunately, this procedure was changed before the final administration of NALS during 1992. Instead of skipping houses where the person who answered the door could not read or respond, notations were made about the reason for not being able to take the assessment. These individuals were then assigned low scores (not zeros as originally proposed). A summary of some of the information from NALS (NCES, 1993a) is shown in Table 3.

Participation of individuals with disabilities in national data-collection programs is constrained due to the lack of accommodations in the assessments. NAEP, for example, provides no modifications. Similarly, NALS did not provide accommodations or adaptations. It is expected that in the near future it will no longer be considered appropriate to conduct a national assessment without allowing proper accommodations for individuals who need them. Trends in this direction are already evident in a major study that is now being conducted about students who are excluded from NAEP. This study is attempting to gain information not only on what considerations go into the decision to exclude students from an assessment, but also on the kinds of accommodations

TABLE 3
Summary of the Average Proficiency Scores on the National Adult Literacy Survey (NALS) of Individuals with Selected Disabilities

Disability Category	*Prose Literacy*	*Document Literacy*	*Quantitative Literacy*
Learning Disability	207	203	200
Mental Retardation	143	147	117
Speech Disability	216	213	212
Emotional Condition	225	224	215
Hearing Difficulty	243	239	247
Visual Difficulty	217	215	214
Total Population[a]	272	267	271

Note. The information in this table is based on Figure 1.10 (p. 44) in the NALS report (NCES, 1993a).

[a]Total population includes individuals with and without disabilities combined. NALS does not provide separate information for individuals without disabilities.

and adaptations that might be needed to allow excluded students to participate meaningfully in the assessment. This is a large step forward for our national data-collection programs, which currently exclude nearly 50% of students with disabilities. Inclusion in the national data-collection programs as a whole will enable students to be included in national assessments that use performance-based measures.

4. State Data-Collection Programs

As might be expected, states are moving toward the use of performance-based assessments for many of the same reasons that national data-collection programs have done so. The link between assessment and teaching is also a motivating factor, with the assumption being that testing programs that emphasize higher-order cognitive skills will result in teaching that emphasizes these skills (Nickerson, 1989).

The findings about participation of students with disabilities in national assessments are repeated to a large extent in existing traditional statewide assessment programs. But it might be expected that as new forms of assessment are developed, consideration will be given to ways to include students with disabilities in the assessments.

State Efforts

Before looking at the participation of students with disabilities in statewide assessments that use performance-based measures, it is important to ask which states are using performance-based assessments for statewide testing programs. McLaughlin and Warren (1994), in another book in this Mini-Library, have highlighted what a few states are doing in the area of performance assessments. These states are exemplars for others. How many other states currently are using performance-based measures in their statewide assessments is less clear.

Interest in the use of performance assessments was first reflected in a document prepared by the Council of Chief State School Officers (CCSSO, 1991) for presentation to the Secretary's Commission on Achieving Necessary Skills. At the time of that survey, CCSSO found that 40 states "have or are planning one or more of the three forms of performance assessment [performance, portfolio, and enhanced multiple choice] at the statewide level" (p. iii). CCSSO defined these forms as follows:

Performance—direct demonstration of target skills.

Portfolio—student work accumulated in a folder.

Enhanced multiple choice—analysis of problems using enhanced multiple-choice answers.

The CCSSO report did not indicate which states were already using performance assessments and which were still in some stage of development.

A more recent source of information on the use of performance-based assessments is a survey conducted by the North Central Regional Education Laboratory (NCREL) in collaboration with the Council of Chief State School Officers. Information from this survey is available in complete form on computer disks (NCREL, 1993); recently some of this information was presented in a document titled *Testing in America's Schools*, published by the ETS Policy Information Center (1994). In this document, it was reported that 38 states are using or considering using some form of nontraditional items in their statewide testing programs. The categories of nontraditional items included in this document were

- Enhanced multiple-choice.
- Short-answer open-ended.
- Extended-response open-ended.
- Interview.
- Observation.
- Individual performance assessment.
- Group performance assessment.
- Portfolio or learning record.
- Project, exhibition, demonstration.
- Other.

Definitions of these terms were not given; states used their own interpretation of the terms when providing information on their statewide assessments. States also provided information on the status of their nontraditional items. In some cases, states were still developing the items; in other cases the items were ready to use. The number of states at the various points of development often add up to more than the number of states indicating that they are using each type of assessment. For example, 22 states indicated that they were using or developing extended-response open-ended items, but 16 reported that these items were "ready to use," 7 reported that they were "piloted, being refined,"

and 5 indicated that they had "begun or completed development." These numbers add up to 28 states.

In response to the NCREL survey, states also indicated the content areas in which nontraditional assessments were being used. Writing was the area most frequently mentioned, with 35 states indicating that they had nontraditional items in this area. Writing was followed by math (29 states), reading (21 states), science (18 states), and social studies (14 states). Other areas (e.g., health, history, geography, music) were mentioned by fewer than 10 states each.

Special Education in Statewide Performance Assessments

In some states, reporting the number of students with disabilities who are excluded or exempted from the statewide assessment is a required part of the accountability system, and when the percentage of exclusions is too high, follow-up monitoring of the appropriateness of exclusions occurs.

In the past, states have been asked by the National Center on Educational Outcomes (NCEO) about the participation of students with disabilities in their statewide assessments (see Shriner & Thurlow, 1992; Shriner, Thurlow, Gilman, & Tundidor, 1993; Shriner, Spande, & Thurlow, 1994). Over the years that these surveys have been conducted, increasing attention has been paid to documenting the numbers of students with disabilities who participate in statewide assessments. In the first survey, it was found that most states had little idea of the extent to which students with disabilities were included in their statewide assessments, nor did they know whether data on these students could be pulled out separately from the data of other students. In the most recent survey, it is evident that states have started to pay attention to the extent to which students with disabilities participate in the assessments. In some states, reporting the number of students with disabilities who are excluded or exempted from the statewide assessment is a required part of the accountability system, and when the percentage of exclusions is too high, follow-up monitoring of the appropriateness of exclusions occurs. (See Ysseldyke, Thurlow, & Geenen [1994] for additional information on accountability practices.)

It might be expected that because the use of nontraditional assessments in statewide assessments is relatively new, developers would have considered how to include students with disabilities up front as the

items were developed. But a recent survey conducted by the National Center on Educational Outcomes (reported by Shriner, Spande, & Thurlow, 1994) suggests that this is not the case.

NCEO researchers contacted the assessment person in each of the states that reported the use of nontraditional items to NCREL. Of the 30 states that had indicated to NCREL that they were currently using or pilot testing nontraditional items, only 21 states indicated, in response to the NCEO survey, that they currently were using or pilot testing nontraditional items (four states did not respond to the NCEO survey). Information on the content areas in which states were using nontraditional items, and the types of items being used, are presented in Table 4.

It seems that most states have not been forward thinking about the inclusion of students with disabilities in their development of nontraditional items for statewide assessments.

When asked to indicate the number of students with disabilities who had participated in these nontraditional assessments, only seven states were able to report a number; another two states could give an estimated percentage of students with disabilities. And, of these nine states, only two were able to break their information down by category of disability. It seems that most states have not been forward thinking about the inclusion of students with disabilities in their development of nontraditional items for statewide assessments.

Accommodations and Adaptations in Statewide Performance Assessments

It also might be expected that states beginning to use nontraditional items in their statewide assessments might be more careful in their planning of accommodations and adaptations used during the assessments. Examples of accommodations and adaptations include using a braille version of an assessment (a modification in presentation format); letting a student give answers orally rather than on a test form (a modification in response format); giving a student more time to complete an assessment (a modification of time/scheduling); and having a student take an assessment in a carrel instead of in a large room with many other students (a modification of setting).

When asked by NCEO about their guidelines for accommodations and adaptations, 5 of the 21 responding states indicated that they allowed no accommodations or adaptations in these assessments (see

TABLE 4
Content Areas and Types of Items in Statewide Nontraditional Assessments

Content Area	*Number of States*	*Types of Items*
Writing	17	Enhanced multiple-choice Short-answer open-ended Extended-response open-ended Individual performance assessment Group performance assessment Portfolio or learning record Project, exhibition, demonstration Other, nonspecified
Math	11	Enhanced multiple-choice Short-answer open-ended Extended-response open-ended Observation Individual performance assessment Group performance assessment Portfolio or learning record Other, nonspecified
Reading	9	Enhanced multiple-choice Short-answer open-ended Extended-response open-ended Observation Individual performance assessment Project, exhibition, demonstration Other, nonspecified
Science	3	Enhanced multiple-choice Short-answer open-ended Extended-response open-ended Other, nonspecified
English/ Language Arts	2	Enhanced multiple-choice Short-answer open-ended Extended-response open-ended Other, nonspecified
Other	3	Enhanced multiple-choice Short-answer open-ended Extended-response open-ended Individual performance assessment Other, nonspecified

Note. The information in this table is based on results from the NCEO survey (see Shriner, Spande, & Thurlow, 1994).

Table 5). Another eight states relied on the IEP to delineate the specific accommodations or adaptations allowed for an individual student. In four states, accommodations and adaptations in the areas of presentation format, response format, time or scheduling adjustments, and setting changes were allowed. Two of these states were very broad in their guidelines, indicating either that any modification that is made during instruction is allowed during assessment, or that any modification that ensures inclusion is allowed. In another three states, a subset of these four types of modifications was allowed.

TABLE 5
Types of Accommodations/Adaptations in Nontraditional Assessments

State	*Alternate Presentation*	*Alternate Response*	*Time/Schedule Adjustments*	*Setting Changes*	*Individualize via IEP*	*None*
Alabama					X	
Alaska						X
Arizona	X	X	X	X		
Arkansas						X
California						X
Connecticut*						
Delaware*						
Florida					X	
Georgia					X	
Illinois	X					
Kansas						X
Maryland	X	X	X	X		
Minnesota					X	
Nevada	X				X	
New Jersey					X	
New York					X	
Ohio					X	
Oregon	X	X	X			
Pennsylvania*	X	X	X	X		
Tennessee						X
Vermont*	X	X	X	X		

*These states gave unique responses. Connecticut indicated that it is in the process of revising its guidelines. Delaware indicated that a range of accommodations is used, but did not specify what they are. Pennsylvania indicated that it is allowing anything that ensures inclusion. Vermont allows anything that is allowed in instruction. The latter two were coded as allowing the four types of accommodations/adaptations because no specification was given that the IEP had to list what was allowed.

5. Conclusions

It is possible to draw some general conclusions about the role of performance assessments in national and state data-collection programs and the extent to which students with disabilities are included in these assessments. The critical issues to address are whether students with disabilities should be included in these assessments, and if they are included, what types of modification should be allowed to increase their participation.

There is some evidence that the use of performance assessments may not benefit students with disabilities. For example, Baker, O'Neil, and Linn (1991) reported that there were differences in the rates at which students attempt more open-ended items:

> The NAEP finding raises equity concerns for the widespread use of these assessments in high-stakes roles, particularly because students in disadvantaged classrooms may have relatively few instructional experiences demanding complex performance over extended time. (p. 16)

Fulford (1991) noted, on the other hand, that equity and fairness are issues that need to be addressed. She also argued that alternative assessments hold the *possibility* of more equitable student measurement:

> These "tests" could accommodate for individual differences with their flexible design and multiple, instead of single checks. Unlike standardized tests, they can account for students' different learning styles and skills, and can measure students' ability to reason and problem-solve in authentic situations. (p. 7)

These varying opinions highlight the fact that (a) a wide array of stimulus and response requirements is lumped into the term *performance assessment*, and (b) little research has been conducted on performance assessment of any type (see Elliott [1994] in this Mini-Library). The use

of performance assessments in national data-collection programs has been relatively narrow in scope (focusing mainly on short-answer written responses). States, for the most part, have also used somewhat limited forms of performance assessments, although states do tend to be more willing to try more extended tasks and formats such as using portfolios and demonstrations.

FairTest (1990), a national organization to promote fair and open testing, has identified the following unresolved issues related to performance assessments:

> First, questions of potential race, class, culture, and gender biases in the new assessments have only begun to be addressed.
>
> Second, the relationship between classroom-based assessments, such as portfolios, and externally administered tests . . . has not been resolved. . . .
>
> Third, simply labeling a test "performance-based" does not make it a good test. . . .
>
> Fourth, states and other agencies must do further work on technical problems to ensure that performance-based exams validly and reliably cover content areas; serve as tests worth teaching to; are not corrupted in the way that teaching to multiple-choice tests has corrupted both test results and instruction; and provide aggregatable data for state and national information.
>
> Finally, meaningful assessments cannot be meaningfully implemented without changes in curriculum, instruction, and school structures. (p. 14)

The Council of Chief State School Officers (1993b) also identified equity as an issue that must be addressed in performance assessments, particularly in relation to authenticity and the observation that authenticity can lead to inequity when "tasks are within the experience of certain populations and not others" (p. 8). CCSSO gave the following specific example that recognizes the complications of disability: "Asking students to write about learning a sport, which is biased against those students whose disabilities, geographic location, or economic status have prevented [them] from learning a sport" (p. 8).

The belief that with new forms of assessment students with disabilities could be included from the start is largely unsubstantiated by the data.

All of this is occurring within the context of national and state assessment programs that either (a) do not know how many students with disabilities participate in the assessments or (b) exclude large percentages of students who could participate in the assessments. The belief that with new forms of assessment (e.g., performance assessment) students with disabilities could be included from the start (i.e., during the development phase) is largely unsubstantiated by the data. The only conclusion that can be reached is that assessment programs that have been inclusive of students with disabilities in the past (i.e., in traditional assessments) tend to be inclusive of students in performance assessments.

There are many ways to promote the participation of students with disabilities in large-scale assessments. Key aspects of doing so will include the following:

- Clarification of guidelines for exclusion/inclusion, covering guidelines related to test development, testing, and reporting of results.
- Use of reasonable accommodations, adaptations, and other modifications in assessment procedures (i.e., ones that would not threaten the technical adequacy of an assessment, such as using an interpreter for a student with a significant hearing impairment to give directions that are typically given orally).
- Monitoring of participation levels.
- Research on the effects of various modifications in assessments (including the use of different types of performance assessments) on the performance of students with disabilities and on the technical characteristics of the instruments.

Clearly, performance assessments are here for good reasons. However, the dramatic increase in the use of traditional assessments in the 1950s and 1960s also occurred for valid reasons. There is a need to conduct research on performance assessments that are to be used within large-scale assessments (both national and state) in terms of both the purpose of the assessment (public information, program improvement,

individual performance) and the effects of the assessment (type of diploma student receives, receipt of school financial incentives, changes in instruction, modification of curricular frameworks). To date, the use of performance assessments has not increased the participation of students with disabilities.

References

Baker, E. L., O'Neil, H. F., & Linn, R. L. (1991, August). *Policy and validity prospects for performance-based assessments.* Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Berlak, H., Newmann, F. M., Adams, E., Archibald, D.A., Burgess, T., Raven, J., & Romberg, T. A. (1992). *Toward a new science of educational testing and assessment.* Albany: State University of New York.

Burton, G., Clements, D., Coburn, T., Delgrande, J., Firkins, F., Joyner, J., Leiva, M. A., Lindquist, M. M., & Morrow, L. (1991). *First grade book: Curriculum and evaluation standards for school mathematics addenda series grades K–6.* Reston, VA: National Council of Teachers of Mathematics.

Cannell, J. J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice, 7*(2), 5–9.

Council of Chief State School Officers. (1991). *Performance assessments in the states.* Washington, DC: Pelavin Associates.

Council of Chief State School Officers. (1993a). *Arts education assessment framework.* [Draft]. Washington, DC.

Council of Chief State School Officers. (1993b). State education assessment center report: Using performance assessment raises issues. *Concerns, 40,* 7–9.

Education Commission of the States. (1992, Spring). Survey of national assessment initiatives. *State Education Leader,* pp. 8–10.

Elliott, S. N. (1994). *Creating meaningful performance assessments: Fundamental concepts.* Reston, VA: The Council for Exceptional Children.

ETS Policy Information Center. (1994). *Testing in America's schools.* Princeton, NJ: Educational Testing Service.

FairTest. (1990). Performance-based assessment spreads across nation. *Fairtest Examiner, 4*(3), 16, 12–14.

Fulford, N. (1991). *Assessment: More than a tool—A driving force in educational reform* (Policy Briefs Nos. 15 & 16). Oak Brook, IL: North Central Regional Educational Laboratory.

Haney, W. (1991). We must take care: Fitting assessments to functions. In V. Perrone (Ed.), *Expanding student assessment* (pp. 142–163). Alexandria, VA: Association for Supervision and Curriculum Development.

Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. *Phi Delta Kappan, 70*(9), 683–687.

Hill, C., & Larsen, E. (1992) *Testing and assessment in secondary education: A critical review of emerging practices*. Berkeley: National Center for Research in Vocational Education, University of California at Berkeley.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes*. Alexandria, VA: Association for Supervision and Curriculum Development.

McGrew. K. S., Thurlow, M. L., & Spiegel, A. N. (1993). An investigation of the exclusion of students with disabilities in national data collection programs. *Educational Evaluation and Policy Analysis, 15*(3), 339–352.

McLaughlin, M. J., & Warren, S. H. (1994). *Performance assessment and students with disabilities: Usage in outcomes-based accountability systems*. Reston, VA: The Council for Exceptional Children.

Mehrens, W. (1992). Using performance assessments for accountability purposes. *Educational Measurement: Issues and Practice, 11*(1), 3–9, 20.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*(3), 229–258.

Mullis, I. (1990). *The NAEP guide: A description of the content and methods of the 1990–92 assessments*. Washington, DC: National Assessment of Educational Progress.

National Academy of Education. (1992). *Assessing student achievement in the states (The first report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: 1990 Trial State Assessment)*. Stanford, CA: National Academy of Education, Stanford University.

National Center for Education Statistics. (1993a). *Adult literacy in America*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Center for Education Statistics. (1993b). *Can students do mathematical problem solving?* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Center for Education Statistics. (1993c). *NAEP 1992 mathematics report card for the nation and the states*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Center for Education Statistics.. (1993d). *NAEP 1992 reading report card for the nation and the states*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Transforming testing in America*. Boston: Boston College.

National Council on Education Standards and Testing. (1992). *Raising standards for American education*. Washington, DC: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1993, October). *Assessment standards for school mathematics* (Working Draft—Highlights). Reston, VA: Author.

National Research Council. (1993). *National science education standards: July '93 progress report*. Washington, DC: National Committee on Science Education Standards and Assessment.

Newmann, F. (1991). Linking restructuring to authentic student achievement. *Phi Delta Kappan, 72*(6), 458–463.

Nickerson, R. S. (1989). New directions in educational assessment. *Educational Researcher, 18*(9), 3–7.

Office of Technology Assessment, Congress of the United States. (1992). *Testing in American schools: Asking the right questions*. Washington, DC: Author.

North Central Regional Educational Laboratory. (1993). *State student assessment program database, 1992–1993* (brochure). Oak Brook, IL: Author.

Pipho, C. (1985, May). Tracking the reforms, Part 5: Testing—Can it measure the success of the reform movement? *Education Week, 22*, p. 19.

Shriner, J. G., & Thurlow, M. L. (1992). *State special education outcomes 1991*. Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Shriner, J. G., Spande, G. E., & Thurlow, M. L. (1994). *State special education outcomes 1993*. Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Shriner, J. G., Thurlow, M. L., Gilman, C., & Tundidor, M. (1993). *State special education outcomes 1992*. Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Shriner, J. G., Ysseldyke, J. E., & Thurlow, M. L. (1994). Standards for all American students. *Focus on Exceptional Children, 26*(5), 1–19.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46*(7), 41–47.

Ysseldyke, J. E., Thurlow, M. L., & Geenen, K. (1994). *Implementation of alternative methods for making educational accountability decisions for students with disabilities* (Synthesis Report No. 12). Minneapolis: National Center on Educational Outcomes, University of Minnesota.

The ERIC/OSEP Special Project

The ERIC/OSEP Special Project at The Council for Exceptional Children facilitates communication among researchers sponsored by the Office of Special Education Programs (OSEP) in the U.S. Department of Education, and it disseminates information about special education research to audiences involved in the development and delivery of special education services. These audiences include

- Teachers and related services professionals.
- Teacher trainers.
- Administrators.
- Policymakers.
- Researchers.

The activities of the ERIC/OSEP Special Project include tracking current research, planning and coordinating research conferences, and developing a variety of publications that synthesize or summarize recent research on critical issues and topics. Each year, the Special Project hosts a conference attended by research project directors sponsored by OSEP. Throughout the year, it holds research forums and work groups to bring together experts on emerging topics of interest. Focus groups representing the Special Project's audiences are held to inform both OSEP and the Special Project of audience information needs and to enhance the utility of publications produced by the Special Project. These publications include an annual directory of research projects as well as publications about current research efforts.

The ERIC/OSEP Special Project is funded under a three-party contract between The Council for Exceptional Children, the Office of Special Education Programs, and the Office of Educational Research and Improvement, U. S. Department of Education. Under this contract, OSEP funds the ERIC/OSEP Special Project, and OERI funds the ERIC Clearinghouse on Disabilities and Gifted Education. The ERIC Clearinghouse on Disabilities and Gifted Education is one of 16 clearinghouses of the Educational Resources Information Center (ERIC) system, which maintains a database of over 440,000 journal annotations and 340,000 document abstracts concerning education. The ERIC Clearinghouse on Disabilities and Gifted Education gathers and disseminates information on all disabilities and on giftedness across age levels.

A product of the ERIC/OSEP Special Project
The ERIC Clearinghouse on Disabilities and Gifted Education
The Council for Exceptional Children

ISBN 0-86586-251-6